CRT-2000-02

Centre de recherche sur les transports (C.R.T.)
Centre for Research on Transportation

ON GENERALIZED MINIMUM SPANNING TREES

by

Corinne Feremans
Martine Labbé
Gilbert Laporte

January 2000

100-05358/

PUBLICATION
CRT-2000-02

*ON GENERALIZED MINIMUM SPANNING TREES**

by

Corinne Feremans[1]
Martine Labbé[1]
Gilbert Laporte[2]

* *This work was partially supported by the Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture, by the Région Bruxelles-Capitale and by the Canadian Natural Sciences and Engineering Research Council (NSERC). This support is gratefully acknowledged.*

[1] Institut de Statistique et de Recherche Opérationnelle, Service de Mathématiques de la Gestion, CP 210/01, Université Libre de Bruxelles, boul. du Triomphe, B-1050 Bruxelles, BELGIUM <cfereman@smg.ulb.ac.be> <mlabbe@smg.ulb.ac.be>

[2] École des Hautes Études Commerciales de Montréal, 3000, chemin de la Côte-Sainte-Catherine, Montréal, Canada H3T 2A7 et Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, CANADA H3C 3J7 <gilbert@crt.umontreal.ca>

Dépôt légal – Bibliothèque nationale du Québec,
Bibliothèque nationale du Canada, 2000

ABSTRACT

The Generalized Minimum Spanning Tree Problem (GMSTP) consists of designing a minimum cost tree spanning several clusters. The purpose of this note is to pinpoint several inaccuracies contained in a previous publication and to propose a valid formulation for this problem.

Keywords: Generalized Minimum Spanning Tree Problem.

RÉSUMÉ

Le problème de l'arbre de recouvrement généralisé consiste à déterminer un arbre de recouvrement de coût minimal incident à plusieurs ensembles de sommets. L'objectif de cette note est de souligner plusieurs imprécisions contenues dans une publication récente et de proposer une formulation valide pour ce problème.

Mots-clés: Problème de l'arbre de recouvrement généralisé.

1 Introduction

The *Generalized Minimum Spanning Tree Problem* (GMSTP) is defined on an undirected graph $G = (V, E)$, where $V = \{1, \cdots, n\}$ is a vertex set, $E = \{(i, j) : i, j \in V, i < j\}$ is an edge set, and V is a union of K *clusters* V_k $(k = 1, \cdots, K)$. A non-negative cost matrix $C = (c_{ij})$ is associated with E. The GMSTP consists of designing a minimum cost tree containing *at least* one vertex from each cluster. Using the same convention as Fischetti, Salazar and Toth (1995), if the tree must contain *exactly* one vertex per cluster, the problem is referred to as the EGMSTP. As far as we are aware, the EGMSTP was first studied by Myung, Lee and Tcha (1995). These authors have shown that the EGMSTP is NP-hard and that unless $P = NP$, no polynomial-time heuristic having a bounded worst-case performance ratio can exist for this problem. They have also proposed a number of integer linear

programming formulations for the EGMSTP, as well as a heuristic and a branch-and-bound algorithm. Recently, Feremans, Labbé and Laporte (2000) have performed a polyhedral analysis of the problem. Dror, Haouari and Chaouchi (2000) have proved that the GMSTP is NP-hard and have proposed several formulations and heuristics for this problem.

The purpose of this note is to show that some of the results proved by Dror et al. derive directly from those of Myung et al. and that three of their formulations are incorrect. We conclude by proposing a valid formulation for the GMSTP.

2 Complexity results

When C satisfies the triangle inequality, there always exists an optimal GMSTP solution containing exactly one vertex per cluster. Thus the GMSTP and the EGMSTP have the same solution in this case. Put differently, the EGMSTP is a particular case of the GMSTP and the complexity results derived by Myung et al. for the EGMSTP apply directly to the GMSTP. Dror et al. have shown that none of their four polynomial-time heuristics has a bounded worst-case performance. This is a direct consequence of the result proved by Myung et al.

3 Formulations

Dror et al. first propose a network flow formulation for the GMSTP obtained by replacing each edge by two opposite arcs. While the formulation is correct, the Lagrangean problem derived from it is not. Indeed the objective function defined by (8) should read:

$$\text{minimize} \sum_{(i,j)\in U'} c_{ij}x_{ij} + \sum_{k=1}^{K} \sum_{(i,j)\in U'} \lambda_{ijk}(y_{ijk} - x_{ij}).$$

In other words, x has only two indices and the first summation should not be over k.

The connectivity-based formulation is also incorrect. Indeed, it contains the constraints

$$\sum_{e \in \delta(S)} x_e \geq 1 \quad (\forall S \subset V, V_k \subseteq S, k \in \{1, \cdots, K\}),$$

where x_e is a binary variable equal to 1 if and only if edge $e \in E$ belongs to the solution, and $\delta(S) = \{(i,j) \in E : i \in S, j \notin S \text{ or } i \notin S, j \in S\}$. These constraints may in fact eliminate all optimal solutions. For example, suppose that $V = \{1,2,3\}$, $V_1 = \{1\}$, $V_2 = \{2,3\}$, $c_{1,2} = 1$, $c_{1,3} = 2$, $c_{2,3} = 3$, so that the unique GMSTP optimal solution consists of the edge $(1,2)$. Then imposing $\sum_{e \in \delta(S)} x_e \geq 1$ for $S = \{1,2\}$ forces edge $(1,3)$ or edge $(2,3)$ to belong to the solution.

Similarly, formulation IP_D is too strong since it contains the constraints $\sum_{e \in \delta(V_k)} x_e = 1 \quad (k = 1, \cdots, K)$ which force each cluster to have exactly one edge connecting it to its complement. This is clearly not valid for the GMSTP.

4 A valid GMSTP formulation

A valid formulation for the GMSTP is the following. Let x_{ij} be a binary variable equal to 1 if and only if edge (i,j) belongs to the solution, and let y_i be a binary variable equal to 1 if and only if vertex i belongs to the solution. Then the GMSTP is to

$$\text{minimize} \sum_{i<j} c_{ij} x_{ij} \tag{1}$$

subject to

$$\sum_{i \in V_k} y_i \geq 1 \qquad (k = 1, \cdots, K), \tag{2}$$

$$\sum_{i,j \in S, i<j} x_{ij} \leq \sum_{i \in S \setminus \{v\}} y_i \qquad (v \in S \subseteq V, S \neq \emptyset), \tag{3}$$

$$\sum_{i,j \in V, i<j} x_{ij} = \sum_{i \in V} y_i - 1, \tag{4}$$

$$x_{ij} \in \{0,1\} \qquad (i,j) \in E, \tag{5}$$

$$y_i \in \{0,1\} \qquad i \in V. \tag{6}$$

Here constraints (2) ensure that the tree spans all the clusters, constraint (3) prevent the formation of cycles, and constraint (4) forces the solution to have as many edges as vertices, minus one.

Aknowledgements. This work was partially supported by the Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture, by the Région Bruxelles-Capitale and by the Canadian Natural Sciences and Engineering Research Council under grant OGP0039682. This support is gratefully acknowledged.

References

[1] M. Dror, M. Haouari, J. Chaouachi. *Generalized Spanning Trees.* European Journal of Operational Research **120** (2000) 583–592.

[2] C. Feremans, M. Labbé, G. Laporte. *Polyhedral Analysis of the Generalized Minimum Spanning Tree Problem.* Technical Report IS-MG 2000/01, Université Libre de Bruxelles, CP 210/01, B-1050 Bruxelles, Belgique. (2000). Available at **http://smg.ulb.ac.be** (Preprints).

[3] M. Fischetti, J.J. Salazar, P. Toth. *The Symmetric Generalized Traveling Salesman Polytope.* Networks **26** (1995) 113–123.

[4] Y.S. Myung, C.H. Lee, D.W. Tcha. *On the Generalized Minimum Spanning Tree Problem.* Networks **26** (1995) 231–241.